**UNITED STATES DEPARTMENT OF COMMERCE**
**Bureau of the Census**
Washington, D.C. 20233

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-1

September 27, 2000

**MEMORANDUM FOR:** Executive Steering Committee for ~~A.C.E.~~ E. Policy ("ESCAP")

Through: John H. Thompson
Associate Director for Decennial Census

From: Howard Hogan
Chief, Decennial Statistical Studies Division

Subject: Data and Analysis to Inform the ESCAP Recommendation

The purpose of this document is to summarize and synthesize the accompanying analysis reports and thereby enable the ESCAP Committee to make informed determinations about the quality of both the initial census and the A.C.E. This document will not make a recommendation on the ultimate question of whether the ESCAP should recommend in favor of reporting the statistically adjusted data as the P.L. 94-171 data.

The list of planned analysis reports was prepared in consultation with the ESCAP Committee and is intended to encompass the initial information that should inform the ESCAP recommendation. No one paper is determinative; rather the information in the analysis reports, taken together, should permit the Committee to evaluate the quality of both the census and the A.C.E. The information in the analysis reports, and the reports themselves in draft form, will be shared with the Committee on a flow basis so that the Committee can evaluate the data as they become available. I invite the Committee to ask for additional information.

The authors of the attached analysis reports and the ESCAP Committee as a whole will analyze the best data available at the time the decision must be made. The ESCAP should consider all useful data available at the time, including data not specified in this memorandum. Similarly, if some of the hoped-for data or analysis is not available, ESCAP should make its recommendation based upon the totality of the available data, taking into account the possible impact of the unmeasured aspect.

# Data and Analysis to Inform the ESCAP Recommendation

## 1. Introduction

### 1.1 Background

As part of Census 2000 the Census Bureau is conducting the Accuracy and Coverage Evaluation (A.C.E.), an operation designed to permit correction of the initial census results to account for missed individuals and erroneous enumerations. The Census Bureau has made a preliminary determination that the A.C.E., if properly conducted, should produce more accurate census data by improving coverage and reducing differential undercounts, but this initial determination is subject to evaluation after the numbers are produced. The Census Act (P.L. 94-171) requires the Census Bureau to release tabulated data to state redistricting officials no later than April 1, 2001. Accordingly, the Census Bureau must decide in early 2001 whether to release statistically corrected data as its P.L. 94-171 data. If the statistically corrected data are released as the P.L. 94-171 data, the unadjusted data will also be made available. A proposed regulation would require the Executive Steering Committee on A.C.E. Policy (ESCAP) to prepare a written report to the Director of the Census recommending the methodology to be used in making the tabulations of population reported to States and localities pursuant to P.L. 94-171. The Director would then be charged with making the decision on whether to release the corrected data.

The purpose of this document, and of the accompanying analysis reports prepared by Census Bureau staff, is to inform the ESCAP Committee with facts and analysis relating to the quality of both the initial census and the A.C.E. This document attempts to summarize and synthesize the analysis reports. The attached documents are prototypes of the analysis reports that we will prepare for the ESCAP Committee upon completion of the operational phase of the census. Each examines an aspect of census or A.C.E. quality. Many of the analysis reports are primarily empty shell tables with descriptive written analysis. Each author will, after completion of the applicable operation or operations, compile the data to fill in the shell tables and will prepare a written analysis of the data. This document will then attempt to summarize and synthesize all of the analysis reports.

The list of planned analysis reports was prepared in consultation with the ESCAP Committee and is intended to encompass the initial information that should inform the ESCAP decision. No one paper is determinative; rather the information in the analysis reports, taken together, should permit the Committee to evaluate the quality of both the census and the A.C.E. It is important to note, however, that the conduct or results of either the census or the A.C.E. may require the authors of these reports to modify their prespecified format, either by including additional information the author determines to be relevant, or by excluding information that was

unavailable or uninformative.

The ESCAP Committee will prepare a written recommendation to the Director regarding the suitability of using statistically corrected data for redistricting. The information in the analysis reports, and the reports themselves in draft form, will be shared with the Committee on a flow basis so that the Committee can evaluate the data as they become available. The Committee may ask for additional information, either from the authors of the analysis reports or from other Census Bureau staff. When the ESCAP Committee makes its recommendation to the Director, it will release publicly all of the documentation underlying its recommendation, including the analysis reports.

Much of the analysis in the attached reports is applicable to evaluating the overall quality of the initial census and the A.C.E. data for all purposes. In some instances, the analysis reports will focus on the ESCAP Committee's regulatory charge: to make a recommendation on the suitability of using the A.C.E. data for redistricting, rather than focusing on federal funds distribution or any other possible use of the corrected data.

### 1.2 Document Structure

*Analysis Reports Important to this Section:*

♦ Document No. 2, "Overall Census and A.C.E. Quality Indicators"

This document will synthesize the attached analysis reports through the following structure:

- Section 2: Review of the Quality of the Uncorrected Census
- Section 3: Review A.C.E. Operations
- Section 4: Review of A.C.E. Quality
- Section 5: Assessing the Service Based Enumeration Estimation Results
- Section 6: Forming an Overall Assessment

This synthesis, when final, should summarize and analyze the available data as a whole, rather than focusing on any single analysis report or any one piece of datum. This document will not, however, make a recommendation or determination on the ultimate question: whether the ESCAP should recommend in favor of reporting the statistically adjusted data as the P.L. 94-171 data. The ESCAP Committee, deliberating as a whole, will compare the overall quality of the census and the A.C.E. to determine, in light of historical experience, whether the statistically corrected data are superior for redistricting purposes.

The authors of the attached analysis reports and the ESCAP Committee as a whole will analyze the best data available at the time. The analysis reports represent the ESCAP Committee's current thinking about what information will be both available and useful. It is important to remember, however, that should other useful data present themselves or become necessary, the

ESCAP will not blind itself to these new data, simply because they were not pre-specified. Similarly, if some of the hoped-for data or analysis is not available, the Census Bureau will make its decision based upon the totality of the available data, taking into account, of course, the possible impact of the unmeasured aspect.

Finally, it should be noted that, as in past censuses, the Census Bureau will prepare in the years following Census 2000 an extensive array of detailed evaluations of many aspects of both the initial census and the A.C.E. These evaluations will not be available until many months after the Census Bureau is required by law to act. These final evaluations, as distinguished from the analytical reports that will inform the ESCAP Committee, will be accomplished without the pressure of a legal deadline, will be based on additional information, and may in some instances reach conclusions different from those in certain analysis reports.

It is also worth noting that Census 2000 has been subject to an unprecedented level of external review, both the initial census and the A.C.E. The ESCAP Committee may also be informed by some external reports from, e.g., the General Accounting Office, the Office of the Inspector General, the Census Monitoring Board, and other sources of information. Although not all of these sources will be of equal quality or relevance, they may provide additional information about the conduct of both the initial census and the A.C.E. They may point to global concerns or to special or local problems. We will review these sources to see if any documentation from outside the Census Bureau can help us assess the relative quality of the initial census and the A.C.E.

## 2. Review of the Quality of the Uncorrected Census

The ESCAP Committee will assess the quality of the A.C.E. in comparison to the quality of the unadjusted census. The Director has publicly announced on several occasions that in deciding which set of numbers to release as the P.L. 94-171 data, that he will informally create a mental two-by-two matrix. This matrix would assess the quality of both the initial census and the A.C.E., with good census/bad census on one axis and good A.C.E./bad A.C.E on the other axis. If, for example, preliminary indications pointed to an initial census of exceptional quality and an A.C.E. of questionable quality, there would be no reason to use the A.C.E. results. Using the A.C.E. to correct a close-to-perfect census would at best only introduce sampling error into the data. If, on the other hand, the initial census results indicated a high net undercount or a high differential undercount in the initial census, along with reliable A.C.E. results, the Director has indicated he would decide in favor of adjustment.

Historical experience does not lead the Census Bureau to expect a close-to-perfect initial census. As documented extensively by Census Bureau and outside statisticians, every census since at least 1940 has experienced both a net undercount and a substantial differential undercount. In particular, the data reveal a persistent differential undercount between the Black and non-Black populations, as well as differential undercounts for other minority groups and for children. The

Census Bureau designed the A.C.E. in the expectation that Census 2000 would continue the historic phenomenon of the differential undercount.

The Census 2000 plan, however, includes several important innovations to the census process that attempt to improve census accuracy. The paid advertising campaign, the simplified census questionnaires, the state-of-the-art mailing strategy, the Partnership Program, the Local Update of Census Addresses program, the New Construction Program, and the Be-Counted program were all designed to produce the best initial census possible, perhaps even an initial census that would not require statistical adjustment. The first step of the ESCAP's decision making process, then, will be to review the results of the initial census to determine whether improved census operations, or societal changes, or other factors, have resulted in a reduction in the level and pattern of undercount to the point where correction, even if possible, would either not be needed or would only be able to make a small and unimportant improvement.

### 2.1 Comparison with Demographic Analysis and Demographic Estimates.

*Analysis Reports Important to this Section:*

- Document 4: "Demographic Analysis Results"
- Document 16: "Demographic Full Count Review"

Demographic Analysis ("DA") has long provided the standard against which census accuracy is measured. See, e.g., Committee on National Statistics, "Modernizing the U.S. Census" (1995) ("Demographic estimates are the primary means for comparing coverage for censuses over time for the nation as a whole."). Indeed, when people discuss the "steady improvement in census accuracy" or say the 1990 census was the "first to be less accurate than its predecessor" they are using Demographic Analysis as a benchmark. See, e.g., the "Report to Congress – The Plan for Census 2000" (1997, page i), and Darga, "Sampling and the Census,"(1999, page 14-15).

Demographic Analysis, as the term is usually used, is the construction of an estimate of the "true" population using birth, death, migration and other data sources independent from either the current census or the A.C.E. DA provides independent measures of the net undercount by age, sex, and Black/Non-black. It represents a generally accepted historic data series, although, of course, it is subject to its own limitations and uncertainties. Among DA's important limitations is the lack of an historical data series to independently estimate the Hispanic, Asian, or American Indian populations. In addition, the level of emigration and undocumented immigration must be estimated using indirect methods. These limitations and uncertainty are documented in Robinson (1993), and Himes and Clogg (1992), as well as in the 1990 "D" studies.

Two more recent concerns warrant mention. The Demographic Analysis method requires reconciling the reporting of race in the vital statistics system with race as reported in the census. For example, in the birth registration system the race of the mother and the father are reported,

rather than the race of child. For the first time, Census 2000 allows for the reporting of more than one race. This change may introduce a new consideration into the reconciliation of reported race data.

A second concern arises from the fact that Demographic Analysis provides estimates for the total population (persons living in households and those in group quarters). This level of analysis is suitable and important to evaluate overall census accuracy, but complicates a comparison of the accuracy of the initial census and the accuracy of the A.C.E. This difficulty is caused by the fact that the A.C.E. measures the net coverage of the housing unit population which does not take into account the group quarters population, including college dormitories and prisons.

Traditionally, we have handled this difference by either subtracting the census count of excluded group quarters from the unadjusted census and DA, or by adding it to the A.C.E. This approach implicitly assumes perfect coverage of group quarters in the initial census. In 1990 the PES ("Post-enumeration Survey") excluded only institutions and military group quarters. The numbers from these sources were small and the assumption of good census coverage was probably warranted. The A.C.E., unlike the PES, excludes all varieties of group quarters, including, most importantly, student facilities such as dormitories and fraternity houses. The number of individuals in excluded facilities is larger and the assumption of ignorable coverage error more problematic. Therefore, for some age groups, especially individuals aged 18-29, care must be taken in these comparisons.

One important strength of Demographic Analysis is its ability to measure sex ratios. While inconsistency in reporting racial data may introduce uncertainty into the Demographic Analysis estimates of a specific population group, in many instances the inconsistency will affect both sexes equally, so that the inconsistency's effect on the expected sex ratio should be quite small (this circumstance may not be true for the group quarters population). In 1990, many of the comparisons between the initial census, the PES, and Demographic Analysis centered on the sex ratios.

Demographic Analysis has two important purposes, in that it can provide an independent reading of the quality of both the initial census and the A.C.E. Although both the A.C.E. and Demographic Analysis are subject to error, the sources and patterns of error are sufficiently different that agreement or disagreement between the two sets of estimates will be critical to the understanding of the A.C.E. results. Specifically, do the areas of disagreement relate to what is already known about the A.C.E., or do they open up new questions and concerns? Do they indicate that the individual errors in the A.C.E. compound each other, or do those errors tend to cancel each other out to permit the formation of useful estimates?

Besides the historic Demographic Analysis data series, the Census Bureau has demographic estimates from the 1990 census for groups such as Hispanics, Asians, Pacific Islanders and the Alaskan Native/American Indian population. Demographic benchmarks are also available for subnational areas. Adjustments can be made in these series for the undercount in 1990 as

measured by the PES. This data set allows us to see if the A.C.E. has produced reasonable estimates of the population by group. Differences between projections from the previous census and the current census are known as the "error of closure" and are a standard measure of census accuracy.

Finally, the Census Bureau works in conjunction with state demographers to analyze, understand and reconcile the results of the initial census with local information on population size and growth. This program, known as the Demographic Full Count Review, will be analyzed and described more fully in Document 16. Its primary purpose is to rapidly examine, rectify if possible, and clear Census 2000 files and products for subsequent processing or for release to the public, but it can also provide information to aid in the understanding of state and local census coverage.

An important analysis will be comparing both the initial and the corrected census measures with demographic benchmarks to determine whether the A.C.E. results generally fit what is expected.

## 2.2 Direct Measures of Census Quality

*Analysis Reports Important to this Section:*

- Document 2: "Overall Census and A.C.E. Quality Indicators"
- Document 3: "Quality of Census Processes"

Demographic Analysis should provide a high level overview of the accuracy of the initial census, as well as important comparisons to historic levels and patterns of coverage. However, all censuses are designed and perform differently, and it is important to also evaluate the strengths and weaknesses of the initial census by reviewing quantitative and qualitative measures of quality. In reviewing the initial census operations, the Census Bureau will seek to understand the results of each major census operation and to evaluate the extent to which the operations were under control.

The following major operations had associated Quality Assurance ("QA") programs:

Address list development
Questionnaire delivery
Nonresponse follow-up
Coverage Edit follow-up
Coverage Improvement Follow-up

We will review each of these operations as well as operations without specific QA programs to see how each was implemented and how its results compare to applicable 1990 levels. When QA results are available, these will be provided and reviewed. Our review will be wide-ranging,

but can be expected to address the following questions:

- Were there gaps, duplicates, or incorrectly geo-coded housing units in the housing unit inventory for the initial census, and is our knowledge about the housing unit inventory consistent with A.C.E. results?

- What do we know about the level of person and housing unit duplication in the census? How might this level affect the pattern of net undercount as measured by the census?

- Will the results of large household follow-up, validate the pattern of undercount for, say, children as measured by the A.C.E.?

- What do we know about the level and pattern of units with no established population counts, units for which the number of people had to be statistically estimated?

- What do we know about the level of whole person imputation in the census, cases where there was evidence that a person existed, but all individual characteristics had to be statistically estimated?

The quality of the nonresponse follow-up operation is particularly important. A detailed analysis using the reinterview forms themselves will not yet be available, but we will review the QA analysis to assess the NRFU program to see if its results are consistent with A.C.E. results. For example, the A.C.E. will measure a level and pattern of fabrication. We will evaluate whether that level and pattern are consistent with the results of the NRFU Quality Assurance program.

Census 2000 will also generate a wealth of non-statistical information that may be used to evaluate basic census quality. Routine administrative reporting may document operational problems that are indicative of special and/or local problems. While we have not designated a formal document to describe this possibility, we will review routine administrative records to provide the ESCAP Committee with all available information about the quality of the initial census.

## 3. Review of A.C.E. Operations

Similar to its review of the operations in the initial census, the Census Bureau will review the A.C.E. operations to identify any deviations from specified procedures and to assess the extent to which the operations were "under control."

### 3.1 Proper Execution of the Steps Between Processing and Estimation.

*Analysis Reports Important to this Section:*

- Document 7: "Missing Data Results"
- Document 8: "Decomposition of DSE Components"
- Document 9: "Dual System Estimation Results"

One concern often expressed about statistical correction is the notion that statistical processes could be more subject to manipulation than traditional enumeration processes. The Census Bureau does not believe this notion is well founded and has publicly documented the procedures underlying the A.C.E. to assuage these concerns. The first step in reviewing the A.C.E., then, is to evaluate A.C.E. operations to determine whether the pre-specified procedures were followed. No similar attempt is made in this document with regard to the initial census, as there have not been similar allegations with regard to this first phase of the census.

The attached analysis reports will allow a review of each of the steps in the A.C.E. process, from creation of the A.C.E. micro-records to the computation of the final adjustment factors. This review should to a large extent, allow the computation of the results to be independently verified.

Document Number 8, "Decomposition of DSE Components" will present an accounting of all A.C.E. records. Beginning with records with complete data, both post-stratification variables and enumeration status, the accounting then proceeds step-by-step through each stage of missing data adjustment and sample weighting until the final weighted "matched" results are provided, the results that are the input data for the dual system estimates. This report should allow the reader to see clearly how the final results were derived and to understand the relative effect of the estimation steps on the results.

Document Number 7, "Missing Data Results" will show the effects of individual missing data estimation steps upon the weighted "matching" results. It will allow the reader to drill-down, or disaggregate, the data to a more detailed level.

Document Number 9, "Dual System Estimation Results" will provide detailed DSE computations together with useful "roll-ups," that is aggregating the results over age and sex, minority/non-minority or other useful summations. It will allow the reader to verify how the final coverage correction factors are computed from the input data.

These three documents, taken together, should clearly demonstrate how the final coverage correction factors were derived from the micro-level data and should document whether the pre-specified procedures were followed.

### 3.2 Conduct and Control of the A.C.E. Operations.

*Analysis Reports Important to this Section:*

- Document 5: "Person Interviewing Results"
- Document 6: "Person Matching and Follow-up Results"
- Document 7: "Missing Data Results"

The second aspect to this review is to establish that A.C.E. operations were well conducted and well controlled. Analysis Reports Numbers 5, 6 and 7, "Person Interviewing Results," "Person Matching and Follow-up Results," and "Missing Data Results," taken together, should establish the basic parameters of the operational quality of the A.C.E., as well as establishing how well the A.C.E. design was implemented. To the extent possible we will compare operational measures from the 2000 A.C.E. to corresponding measures from the 1990 PES. Was the conduct and control of the A.C.E. better – or no worse – than in 1990?

Both interviewing results and Quality Assurance results should be available by the time the ESCAP Committee must act. The interviewing results will be standard survey indicators, such as non-interview rates, proxy rates, and refusal rates. This information should be available separately for telephone, personal visit and non-response conversion interviewing.

One important aspect of the review of missing data results will be the distinction between interview rates for households on the date of the A.C.E. interview and on Census Day. The Procedure C treatment of movers used in A.C.E. employs data from both households. However, by design, all whole household outmover interviews will be by proxy.

We will also evaluate the results of the A.C.E. interviewing Quality Assurance program. This Quality Assurance program was designed to prevent A.C.E. fabrications and other major interviewing problems. The analysis should provide the level of A.C.E. potential fabrications detected, and estimate the level of remaining undetected fabrications. Additionally, we will evaluate matching and person follow-up operations. Again, simple operational summaries will provide an overview of the results by individual match codes and how the results differ for important levels of summation.

Since matching has an extensive Quality Assurance program, the results from the matching QA program should measure the final level of matching error or uncertainty. The analysis of the matching system will compare the level of matching error for all matchers to the level of matching error of the Census Bureau's most experienced matching group, the ten permanent matching specialists. This analysis will show the extent to which the Census Bureau was able to

control the quality level of the large number of matching clerks who will do the bulk of the matching.

We will assess follow-up interviewing along with our assessment of the matching process because the purpose of the follow-up is to aid the matchers. The level and timing of the follow-up will be given, together with the Quality Assurance results.

Finally, we review the missing data process. The level of missing data reflects the extent to which the interviewing and matching process together are able to provide input to the Dual system estimatess. Levels and patterns of missing data reflect the extent that the input data processes were successfully conducted.

The above-described analysis should provide a preliminary reading about the operational success of the A.C.E.

## 4. Review of A.C.E. Quality

The review in Section 3 should establish whether the A.C.E. was conducted as designed. This section will take the next step and evaluate the quality of the A.C.E., as implemented.

Our review of A.C.E. quality will have two aspects. First, the Census Bureau will review the available data relating to selected individual components of A.C.E. error. It must be emphasized that the Census Bureau will only be able to review data available prior to the statutorily-mandated deadline. The Director is required by law to release redistricting data to the states prior to April 1, 2001. Some information relating to the quality of the A.C.E. will not be available in time to inform the Director's decision. As mentioned earlier, a complete evaluation series will be published several years after the ESCAP's recommendation, in 2003, and will assess the information more completely. The second part of the A.C.E. quality review will synthesize what is known about the components of error into a few indicators of overall relative quality for both the corrected and the uncorrected measures.

### 4.1 **Individual Components of A.C.E. Quality.**

#### 4.1.1 *Sampling Variance.*

*Analysis Reports Important to this Section:*

- Document Number 9: "Dual System Estimation Results"
- Document Number 11: "Variance Estimates by Size of Geographic Area"

The A.C.E., like all sample surveys, is subject to sampling variance. Sampling variance is one of the most easily measured quality indicators. As in the past, we will measure variance for each

post-stratum, however, changes in post-strata mean that we will not be able to compare the A.C.E.'s variances directly to those in the 1990 PES. We will, however, attempt to combine the A.C.E. variances to produce usefully similar categories.

Variances will also be computed at various levels of political geography: states, typical Congressional districts, and certain cities, towns, counties, and places of a given size. We plan to compare the A.C.E. coefficients of variation ("CVs") to the corresponding 1990 CVs at each applicable level. Our analysis and comparisons will focus on the distributions of the CVs at various geographic levels; we will not be comparing the CV for any particular city, county or other substate entity to that entity's 1990 level. The Census Bureau is required to decide whether the A.C.E. numbers, as a whole, are superior to the unadjusted counts, and considers the CV of any given substate entity to be irrelevant to that determination.

The methods that will be used to measure variance should also include many non-sampling variances. For example, if a particular interviewer or clerk introduces random variation into the process, this random effect will be partially reflected in our variance measures. Because of this fact, most of what follows will focus on systematic measurement errors.

#### 4.1.2 *Consistent Reporting of Census Day Residence*

*Analysis Reports Important to this Section:*

- Document 6: "Person Matching and Follow-up Results"

Proper application of the DSE model requires consistent reporting of Census Day residence between the P and E samples. If a person who was sampled in the P sample reported a different Census Day residence than he/she would have reported in the E sample, then that person could be considered both missed (based on the P sample) and correctly enumerated (based on the E sample), or conversely, both enumerated (based on the P sample) and not-correctly-enumerated (based on the E sample). Since many people fall only into either the P or the E sample, measuring consistent reporting is a challenging task. When a person is in both the P and the E sample, we obviously have consistent reporting between the two systems because we use the same interview for both samples. However, some individuals have two interviews, one in the P Sample and one in the E Sample. For example, for individuals in the P sample, we would use the initial A.C.E. interview to determine their correct Census Day residence. However, if an individual was missed by the A.C.E., but included in the initial census, we would use the A.C.E. follow-up interview to determine Census Day residence. Even for matched people, if the person was duplicated by the census, we might have a different interview at each identified census household. Since these interviews use different survey questionnaires, administered at different times, by different interviewers, and to potentially different respondents, there is a chance that the two questionnaires could determine different "correct" Census Day residences for the same person. Although we would never observe this directly, this difference in the "expected value" can influence the Dual system estimates.

The 1990 Evaluations (P studies) measured the consistency of reporting Census Day address in the PES by comparing the reinterview to the production results. (See P - 4, "Address Misreporting"). One problem in 1990 was the misreporting of Census Day address, with an estimated 0.7 percent of the P-sample being erroneously reported as nonmovers. (See P - 4, "Address Misreporting") The 2000 A.C.E. improves on 1990 PES, in particular because the use of the CAPI instrument vastly improves on the use of the 1990 questionnaire by requiring the interviewer to ask all questions in the interview form.

A comparison of the survey instruments from 1990 and 2000 may shed some light on the consistency of the questions for the P and the E samples. For matched people, clearly the same interview was used to confirm both sides. For non-matches, we will have available the proportion of proxy P sample non-matches and the proportion of E sample proxy erroneous enumerations, especially for those claiming usual residence else where. However, the analysis of these kinds of indicators is far from direct. The interview status can also be compared for the initial CAPI interview and for the follow-up interview for the P-Sample nonmatches. We would look at the number of P-Sample nonmatches that have proxy initial interview and non-proxy follow up and were then removed from the P-Sample because they lived elsewhere on Census Day. Such a study could identify whether respondent type is a measure of quality of reporting Census Day Address, allowing us to examine the proxy rates between the P and E-Sample follow-up interviews.

#### 4.1.3 *Matching Error.*

*Analysis Reports Important to this Section:*

- Document 6: "Person Matching and Follow-up Results"

Matching error refers to assigning the incorrect code to a P-sample record. Matching error can consist of assigning a code of "matched" to a true non-match case, and vice-versa. It can also consist of assigning an unresolved code to a case that has sufficient information. Matching errors can directly influence the final Dual system estimates. Errors in matching will have both a random and a systematic component. The random component will be partially reflected in the overall variance estimates.

Matching error was measured in 1990 via a rematch study (P-7, "Estimates of P-Sample Matching Error from a Rematching Evaluation," P-10, "Measurement of the Census Erroneous Enumeration Clerical Error made in the Assignment of Enumeration Status"). A study of clerical error in the 1990 PES found error in coding matches (P-5a, "Analysis of Fabrications from Evaluation Follow-up Data") and erroneous enumerations (P-6, "Fabrication in the P-sample -- Interviewer Effect"). In 1990, codes were entered into a computer system, but the actual matching and duplicate searches were done using paper. In the A.C.E., we expect matching to be better controlled and more efficient because the clerical matching and quality assurance are fully automated. The automated interactive system will not prevent all matching

error, but should reduce the chances for error significantly.

The 1990 matching system matched both non-movers (within the sample area) and in-movers (who could be coded and matched in any area). The non-mover matching system was largely automated. The mover match system not only included several additional steps (mainly to geographically code the Census Day address), but was also completely clerical. For the A.C.E., all matching will be within the sample area or its surrounding blocks. The system will match both non-movers and out-movers. The system is significantly more automated, with little clerical matching. All clerical matching operations will be conducted at one location. Comparisons to 1990 must take these changes into account.

Other examples of the improvements in coding are:

- Electronic filtering allows searching based on first name, last name, characteristics, and addresses. For example, the system allows searching for all people named George, all people whose last name begins with a H, all people on Elm Street, or everyone between 30 and 40.
- Only particular codes that fit the situation are allowed. For example, only P-sample nonmatch codes can be assigned a P-sample nonmatch after follow-up code.
- The electronic searches for duplicates will reduce the tedious searching through paper lists of census people. The searching in 1990 was limited to printouts in two sorts: last name and household by address. In 2000, the clerks will have the filtering on name, characteristics, and address to help identify duplicates.
- The system monitors whether the matcher has completed all the necessary searches such as looking for duplicates.
- There are built in edits to check for consistent coding. For example, codes that apply to a household are assigned to all people in the household, such as a geographic code.
- The system automatically assigns certain codes, minimizing coding error.
- A code to indicate that the case needs review at the next level of matching is available to the clerical matchers. This code allows them to flag unusual cases to be done by a person with more experience.
- All quality assurance for the clerical matching is automated. The quality assurance cannot be skipped in 2000.
- Clerical matching is centralized at the National Processing Center instead of different groups of matchers in the seven processing offices, as was done in 1990. Forty six Technicians were hired in September 1999 and have been thoroughly trained in the design of the A.C.E. and matching of people and housing units. These Technicians will perform the quality assurance for the clerical matchers. Additionally, ten Analysts are our most experienced matchers. The Analysts will do the quality assurance for the Technicians and handle the most difficult cases.

The main direct information we will have available to assess the matching will be the results of the matching Quality Assurance program. This program will give us information about the level

of error relative to that of our most experienced matching specialists. It should be noted that many of these same individuals participated in the 1990 PES.

### 4.1.4 *A.C.E. Fabrications.*

*Analysis Reports Important to this Section:*

- Document 6: "Person Matching and Follow-up Results"

Inclusion of fictitious people in the Dual system estimates can create a strong bias because fictitious records have little chance of being matched between the P and E samples. Fictitious records, of course, should not be included in either system. Fabrications in the initial census are measured by the E sample (See Section 4.1.7 below). Here we concentrate upon fabrications in the P sample.

In 1990, the level of residual fabrication in the P-sample was measured by three studies evaluating different measures of potential fabrication. The first, a study of interviewer fabrication in the 1990 PES (P-5, "Analysis of P-Sample Fabrications from PES Quality Control Data") evaluated interviewer fabrication detected in the quality control operation (and rectified by the QC operation), as well as fabrication detected in the follow-up operation. The estimated number of fabricated persons remaining, at the national level, after the QC operation was approximately 0.13%.

Another study, using data from the 1990 Evaluation Follow-up, concluded that an additional 0.09% (weighted to the PES unweighted totals, this represents 0.03% of the total sample) of the P-sample follow-up interviews included in the evaluation sample should have been coded as fictitious. (P-5a, "Analysis of Fabrications from Evaluation Follow-up Data"). This evaluation was designed to identify P-sample fabrication not detected by the quality control procedure.

A third study, Project P6: "Fabrication in the P-sample -- Interviewer Effect" compared the nonmatch rates of interviewers working in similar areas while assuming that deviations from this nonmatch rate may have indicated undetected curbstoning. This study used a model to predict nonmatch rates, and showed that overall, between 0.9% and 6.5% of the interviewers were found to have high nonmatch rates, high rates that may have corresponded to dishonesty in their data collection.

To help us evaluate potential fictitious records in the A.C.E., we will have detailed Quality Assurance results documenting the level of detected fabrications in the initial A.C.E. interview, as well as measures of residual fabrication. In addition we will have the results of the Person Follow-up interviewing, which should help us identify whole household P-sample fabrications not detected by the interviewing Quality Assurance program. These sources will form the principal basis to evaluate the level of A.C.E. fabrication.

We will also have one more resource. Fabrication is often highly clustered. An otherwise acceptable interviewer might, for example, suddenly fabricate his or her last assignment. The matching analysts will keep a detailed record of any unusual, or "outlier" cluster. These analysts can request special questions during follow-up or send additional interviewers if they question the integrity of one interviewers' results. These records will provide an additional clue to whether there was substantial, clustered fabrication in the P sample. Analysts have the discretion to remove cases they believe to have been fabricated.

#### 4.1.5 *Missing Data.*

*Analysis Reports Important to this Section:*

- Document 7: "Missing Data Results"

Missing data introduces an additional component of uncertainty in the Dual system estimate results. Missing data can contribute to variance and, if the missing data models are poorly specified, can also contribute to bias and differential bias.

Missing data has three components:

- Whole household non-interviews
- Unresolved match, residence or enumeration status
- Missing post-stratification variables.

This section will focus on the first two components of missing data, whole household non-interviews and unresolved match, residence or enumeration status. The third component of missing data, missing postratification variables, will generally result in correlation bias or synthetic error and will be evaluated in connection with the analysis reports on those topics. Missing post-stratification variables tend to lead to correlation bias or synthetic error because this omission can increase heterogeneity and inconsistent post-stratification between the initial census and the A.C.E. High levels of missing data, particularly for unresolved match, residence or enumeration status, also tend to increase variance. We will not evaluate how this type of missing data by itself increases variance because this component is largely picked up in the way we measure the sampling variance.

The 1990 PES dealt with movers by using Procedure-B. Under Procedure-B, missing data can occur when the interviewer fails to get information from the respondent, in either the initial interview or the follow-up interview, or the missing data can occur during follow up. The 1990 PES had very low rates of initial missing data, but a greater number of unresolved cases in the follow up process. Procedure "B" required geocoding the matching, making it possible that completed "mover" cases could not be used because of ambiguities in the geographic coding. Procedure "B," therefore, resulted in initially low rates of missing data, but was responsible for additional missing data in later processes. While the use of Procedure "C" may result in higher

initial rates of missing data, we believe that the total missing data rates may be somewhat comparable.

The effects of missing data on the 1990 results were studied in two ways. First, the modeled results were compared to the results of further field work on the non-response cases (P-3, "Evaluation of Imputation Methodology for Unresolved Match Status Cases"). The field work largely validated the models. This alone is extremely important work as it clearly demonstrated that some of the extreme missing data adjustments sometimes proposed (e.g. assuming all non-response cases were missed) are not supported by the data. Second, additional 1990 studies considered alternative models to the production models (P-1, "Analysis of Reasonable Imputation Alternatives"). Again, these additional studies tended to show the robustness of the results to reasonable alternatives.

There have been two important changes to the survey since 1990 that might affect missing data rates. First, the level of missing data, at least from the A.C.E. interview, might be higher. In 1990 the Census Bureau only needed to interview the current residents, whereas in Census 2000, we seek information about both the current (A.C.E. Interview Day) residents and the Census Day residents. Based on Dress Rehearsal results, we anticipate a higher non-response rate for data concerning the Census Day residents. On the other hand, Procedure C, which we will use in the A.C.E., has eliminated the need to geographically code the Census Day address of "in-movers," thus eliminating one potential source of missing data.

Second, the A.C.E. will use a different missing data model for unresolved match and residence status. The 1990 model was based on hierarchical logistic regression, while the 2000 model will be a far simpler "Imputation Cell Estimator." The input data and behavioral assumptions between the two models are similar but not identical.

Our assessment of the A.C.E. and our comparison with the 1990 PES will begin by assessing the level and patterns of missing data and will also look at the following:

- the effect of non-interview adjustment on weights,
- the distributions of characteristics used to assign post strata before and after characteristic imputation,
- a comparison of the probabilities assigned for unresolved status among the various imputation cells, and
- following the match status of P sample people from before follow up status to final status.

### 4.1.6 *Balancing Error.*

*Analysis Reports Important to this Section:*

- Document 8: "Decomposition of DSE Components"

Balancing error occurs when the set of correct enumerations records defined and measured in the E sample does not correspond to the set of records against which P sample matching is allowed. An important type of balancing error occurs when the search area, as defined and implemented in the E sample, does not correspond to the search area as defined and implemented in the P sample. The dual system model first determines the number of individuals who are correctly in the initial census (through the E sample) and then the proportion of the true population who are correctly in the census (through the P sample). If the E sample and the P sample use different definitions of "correctly in the census," the model will not work.

Balancing error, especially geographic balancing error, was a major concern in the 1980 PEP. The E sample in 1980 counted a person as being correctly in the census only if he or she was counted in the correct Enumeration District (ED). Enumerations outside the correct ED were considered erroneous. However, the P sample searched several EDs looking for a match. Thus some P-sample people were considered correctly enumerated because they matched to census records that would have been considered erroneous had these records been included in the E sample. This problem was virtually eliminated in the 1990 PES by using identical search areas for non-movers. A small concern remained for movers but did not prove to be a problem. (P-11, "Balancing Error Evaluation").

For the A.C.E. we are using a somewhat more complex balancing design. One minor change is that the search area is somewhat smaller, encompassing only the first ring of housing units around a census block. More importantly, not all cases will be eligible for search, coding and matching in the surrounding ring. Only whole household non-matches and E sample geocoding errors are eligible for surrounding block search. This search area is referred to as "Targeted Extended Search" or TES. Finally, the surrounding block search will be performed on a sample basis.

The goal of extended search, whether targeted or not, is to reduce the variance of the estimators, especially for small estimation cells where geocoding errors will not tend to cancel out. While extended search can occasionally reduce the systematic errors in the estimates, this reduction is not a primary goal and the effect should be small. To assess the effect of TES, we will compare correct enumeration rates and match rates for TES and non-TES cases.

Census geocoding error is identified in the housing unit match, which is a match of the A.C.E. inventory of housing units to a preliminary census inventory defined by the January 2000 Decennial Master Address File (DMAF). Census units not on the January DMAF, i.e., 'adds', that are identified in the person follow-up to be geocoding error will have a probability of

correct enumeration imputed. P-sample whole household non-matches in housing units that were matched to census housing units that are no longer in census, i.e., 'deletes', will also be part of the extended search.

### 4.1.7 *Errors in Measuring Census Erroneous Enumerations.*

*Analysis Reports Important to this Section:*

♦ Document Number 6: "Person Matching and Follow-up Results"

Erroneous enumerations occur in the initial census in the following circumstances:

- when an individual had another residence where he or she should have been counted on Census Day,
- when an entry is fictitious,
- when entries are duplicated,
- when an individual lived in a housing unit subject to geocoding error, and
- when the Census Bureau has insufficient information for matching and follow-up.

Errors in measuring census erroneous enumerations can have a serious and direct impact on the A.C.E. For example, a systematic tendency in the A.C.E. processing to code census fictitious cases ("curbstoned cases") as E-sample "non-interviews" would lead to an incorrect estimate of the number of respondents correctly enumerated in the initial census. A tendency to "give the census the doubt" can result in people who move out before Census Day being coded as correct enumerations. While the overlapping P and E samples will lend considerable robustness to the A.C.E. estimates, however, both systematic and random errors can be expected to occur.

E-sample cases are coded either during the initial matching operation, or are based on information gathered during A.C.E. follow-up. For the A.C.E., we will assess this type of error by the analysis of the matching systems' Quality Assurance results, as well as by using information from A.C.E. follow-up. The Quality Assurance program should indicate any problems the clerks had in assigning enumeration status.

One new feature in Census 2000 is an analysis of the level of census duplication. Duplication can now be measured independently of the A.C.E. and the results compared to what the A.C.E. measured. This analysis is possible because the Census 2000 data capture system accurately captures name and other characteristics for the whole population. Thus, it is now possible to do a wide computer search for people who have been counted more than once, thereby providing a measure of duplicates independent of the A.C.E. Although the methods of measurement differ, (for the census as a whole, we will not use clerical matching), the overall pattern and level of duplication can be checked against that measured by the A.C.E. to help validate the A.C.E. E-sample coding.

The Census Bureau found clerical error in assigning erroneous enumerations in 1990. (P-10, "Measurement of the Census Erroneous Enumeration Clerical Error made in the Assignment of Enumeration Status"). The improvements in Census 2000 clerical matching (described earlier in section 4.1.3) should improve the assignment of erroneous enumerations. The identification of duplicates will be closely monitored to assure the duplicate search is done within the block cluster and in the surrounding blocks for TES clusters. The follow-up interview has been improved to guarantee the interviewer has conducted sufficient searches for people to allow accurate coding of fictitious people.

### 4.1.8 *Correlation Bias.*

*Analysis Reports Important to this Section:*

- Document 12: "Correlation Bias"
- Document 4: "Demographic Analysis Results"

In its purest form, dual system estimation assumes that the chance of being included in the P sample is independent of the chance of being correctly included in the initial census. Although this assumption has proven useful in providing a better estimate of the population, it is, of course, unlikely to be absolutely true. Correlation bias can occur from two sources. First, it can be caused by inherent heterogeneity within the post-strata. It can also arise when the event of being enumerated in the census changes the probability of being included in the A.C.E.

Even within post-strata there may be unobservable sub-groups with differing chances of being included in each system. There is also quite likely some group (of an indeterminate size) whose probability of being included in any survey is so low as to be effectively zero. Correlation bias will tend, therefore, to lead to an underestimate of the population. Dual system estimation will estimate some, but not all, of the people omitted from the initial census.

To measure correlation bias, one would ideally like to have an external measure of "truth." Demographic Analysis, especially demographic sex ratios, provide an external measure that, while not perfect, is useful because it is not subject to many of the limitations of the initial census or the dual system estimates.

Using demographic results, the 1990 studies detected a clear pattern of correlation bias in the 1990 PES (P-13, "Use of Alternative Dual System Estimators to Measure Correlation Bias"). Correlation bias was especially strong for adult African-American males, whom the DSE seems to have underestimated considerably.

Recent criticisms of the 1990 studies seem to point to the fact that these studies underestimated the level of correlation bias in the 1990 PES. This conclusion follows from the fact that, in general, correlation bias tends to lower the estimated population, while other measurement errors tend to raise the estimate. Correlation bias and the other kinds of errors therefore may have

tended to cancel each other out. However, this reasoning applies to comparisons of the 1990 PES estimates to DA population totals. If comparisons are instead made to the DA sex ratios (as was done in the P-13 report), and if the other measurement errors are not very different between males and females, then these other measurement errors should tend to cancel out and have little effect on resulting estimates of correlation bias.

An additional problem is that since Demographic Analysis provides national results, one must model how these errors might distribute themselves by post-strata. Several alternative models have been tried. (P-13, "Use of Alternative Dual System Estimators to Measure Correlation Bias"; Bell, "Using Information from Demographic Analysis in Post-Enumeration Survey Estimation," 1993).

The level of correlation bias in the A.C.E. might be larger than that in the 1990 PES because of the use of Procedure C. Procedure C was designed to reduce the difficulty of matching, and thus matching error, by eliminating mover matching. However, since this procedure calls for the reconstruction of the Census Day household, its use may increase correlation bias. Weighting the out-mover match-rate by the number of in-movers may partially, but probably not completely, compensate for the possible increase in correlation bias. Even among outmovers, those more likely to be enumerated in the initial census may be more likely to be picked up in the A.C.E. interview. Because of this potential correlation, we might overestimate the mover match rate.

Our analysis of correlation bias in the A.C.E. results will be, as in 1990, a comparison with Demographic Analysis estimates. The models used to distribute the errors to the post-strata will be similar to those of 1990.

We will examine records and reports for any indication of causal dependence, that is, any indication that participation (or non-participation) in the initial census directly influenced participation in the A.C.E.. For example, we will look at the number of letters received from households that refused to participate in the A.C.E. because they had already sent in their census form. We will look for reports from the regional offices to see if there is any indication of collusion between the census enumerators and the A.C.E. interviewers.

#### 4.1.9 *Synthetic Bias and Synthetic Variability.*

*Analysis Reports Important to this Section:*

- Document 14: "Synthetic Assumptions"

Synthetic estimation error differs from the other measurement errors discussed in this document because it is not directly related to the accuracy of the dual system estimates themselves, but rather to the distribution of the measured net undercount to local areas and demographic subgroups.

Synthetic error can be usefully divided into synthetic bias and synthetic variance. Each person has a probability of being correctly enumerated in the initial census. If these probabilities were equal within post-strata (homogeneity), then the dual system estimates, together with the synthetic model, would produce unbiased estimates of the true population for small areas and subgroups. However, since the DSE model only assumes that the probability of enumeration is equal, some people would be missed. These misses might tend to cluster in certain small areas, thus producing localized variations in geographic coverage. These local variations are synthetic variance. These variations are not necessarily evidence of synthetic bias, but rather are evidence of the fact that the event of inclusion is correlated at the small area. For example, even if all people had the same chance of being counted in the census, (perfect homogeneity in capture probabilities), we would not be at all surprised that whole households/buildings/block-faces tended to be counted or missed together. So clustering of census omissions is not necessarily evidence of synthetic bias, but may be the result of correlated random events.

Another important difference between synthetic error and other types of A.C.E. error is that local heterogeneity is present in the unadjusted census; this local heterogeneity will affect the quality of census results even before A.C.E. adjustment. While this local heterogeneity is not, strictly speaking, synthetic error, since no synthetic estimation is involved, the effect of local heterogeneity on the accuracy of the population estimates is similar. If local heterogeneity in the initial census is correlated with post-stratification variables, then the DSE/synthetic estimation process can reduce this heterogeneity. To illustrate, if a local office manager or crew leader applied the census procedures in a way that resulted in a locally higher net undercount, then the DSE/synthetic model would not correct for this effect. Fortunately, evaluations of the synthetic assumption will help us understand residual heterogeneity in both the initial and the corrected census.

Evaluations of the synthetic assumption are necessarily indirect. Because the A.C.E. is based on a sample, it may be inefficient at detecting truly local heterogeneity. Attempts at measuring local heterogeneity at the block cluster level suffer from the problem that the A.C.E. is not designed to directly measure the undercount, even for the sample clusters. Targeted extended search and large-block subsampling, for example, both allow matching beyond the sample segments. The A.C.E. is designed to measure undercount at high levels, not at the local level.

However, other data are available for all census areas. Some of these data may be related to the net undercount, although in perhaps complex ways. These data for 2000 include:

- census allocation rates,
- census mail return rates, and
- census substitution rates.

An analysis of these data series may identify local variations, (by city or Local Census Office, for example) not related to the post-stratification. This local variation would indicate the possibility of synthetic error. The analysis must be carried out at a high enough level of

aggregation, certainly no smaller than the tract level, so that the effect of synthetic variance would be small.

Assessments of the 1990 PES were concerned with the accuracy of the synthetic assumption for low levels of geography, such as blocks. Our assessment of the synthetic assumption in the A.C.E. accepts that perfect homogeneity cannot exist at the block level. The Census Bureau's evaluation of synthetic error, therefore, will focus on whether heterogeneity at the local level is so great as to prevent an improvement from using the A.C.E., not on whether the post-strata are absolutely homogeneous.

Finally, it is important to note that there is no generally-accepted method to measure synthetic bias in post-enumeration surveys. Also, as explained in the attached analysis report, while the final document will assess the 2000 results in light of the 1990 data, direct comparison between these evaluations and the 1990 evaluations may not be entirely possible.

#### 4.1.10 *Other Measurement and Technical Errors.*

*Analysis Reports Important to this Section:*

- Document 10: "Consistency of Post-Stratification Variables"

The coverage measurement process is subject to several other kinds of measurement errors that need to be noted, including technical ratio bias, contamination error and inconsistent post-stratification.

Technical ratio bias is well documented in the statistical literature and occurs when the expectation (statistical average) of a ratio differs from the expectation of the numerator divided by the expectation of the denominator. Technical ratio bias in survey estimates is usually not important unless the sample size is small. Usually, a sample size of thirty independent observations is adequate (Cochran, 1963). The Dual system estimates is a ratio estimator and as such is subject to ratio bias. Further, since the Procedure C treatment of movers is also a ratio estimator, this treatment of movers introduces a further ratio bias. The A.C.E. is designed to guard against large ratio bias by requiring a minimum cell size for both the post-stratum and the number of outmovers in the Procedure C estimate. While we do not expect technical ratio bias to be a problem in the A.C.E., as part of our review, we will document the median and minimum cell sizes, using standard theory, to verify that this error has not tainted the A.C.E. results.

Contamination error occurs when the conduct of the coverage measurement survey affects how people react to the initial census in the sample areas. If contamination occurs, the coverage measurement survey may no longer reflect the error for the population as a whole, even if it correctly measures the coverage ratios for the sample areas. Contamination error has affected past coverage measurement surveys. The 1980 coverage measurement study (the PEP) was based on the April Current Population Survey, which had been conducted between Census Day

and the start of non-response follow-up. Evidence pointed to contamination error (See Fay et al., 1988). Prior to the 1998 Dress Rehearsal, contamination error was a major concern. See, e.g., Griffiths, "Results from the 1995 Census Test: The Contamination Study" (1996). The Census Bureau planned to conduct non-response follow-up on a sample basis everywhere except for the Integrated Coverage Measurement ("ICM") sample blocks, where this operation would be conducted on a 100% basis. If there was any sampling bias due to the non-response sampling, this bias could differentially affect the ICM and the non-ICM blocks. An evaluation was conducted, but did not detect any contamination.

With respect to possible contamination, the A.C.E. is, with one exception, quite similar to the 1990 PES. In both surveys, housing unit listing was conducted before census mailout and non-response follow-up. Personal visit interviewing was, in both cases, conducted after the end of nonresponse follow-up, but concurrently with various census coverage follow-up field interviews. There is no indication from 1990 that contamination was a problem.

One possible cause of contamination in the A.C.E. was that approximately one third of the A.C.E. interviews were conducted by telephone concurrent with census non-response follow-up. These cases were restricted to cases that had a completed census questionnaire that provided a telephone number, and excluded units in small multi-units structures, and units without house-number-street-name. Because of the timing of this operation, nearly all the interviews were restricted to census mail-back cases. It is possible that some of these cases might have been visited later during non-response follow-up, and that their responses to that operation were influenced by the A.C.E. interview.

While this type of contamination can occur, we believe its effect will be small and difficult to detect. Indeed, as noted above, we have not been able to detect serious contamination since moving away from the 1980 design. The ESCAP analysis of possible contamination will be restricted to reviewing any available reports from the field, the matching clerks or the public that might indicate more than a few people were contacted in advance of their non-response follow-up visit.

Finally, we turn to inconsistency in post-stratification between the A.C.E. and the census. Some individuals may be classified in the initial census into different post-strata than they would in the P sample. The initial census will certainly misclassify some individuals, causing them to be included in the "wrong" category. For example, some Hispanics may be classified as non-Hispanic, or some "American Indians" as "White." To the extent that the true coverage probabilities are equal only for the true characteristics, census mis-classification (i.e. incorrect post-stratification) may introduce correlation bias and synthetic error. These issues are dealt with in those sections.

For the purpose of understanding this possible problem, one should consider the initial census classification as "truth" because, ultimately, the coverage correction factors will be applied to the census person records. If we then take the census classification as the truth, we may measure the

approximate magnitude of the effect of inconsistent post-stratification if we are willing to make one important assumption. We must assume that the level and direction of misclassification of those people in the P sample alone is similar to that of those in both the census and the P sample.

The Census Bureau has never detected a significant impact of classification error on the coverage estimates from any earlier coverage measurement surveys. However, the introduction of multiple race reporting in both the census and the A.C.E. has raised concerns about this type of error.

The impact of inconsistent post-stratification is a function of the proportion of misclassified records and the differences in coverage rates between the two post-strata. If only a few records are inconsistently classified, there will be little impact. Further, there is little impact on coverage if the misclassifications occur between post-strata with similar census coverage rates. Misclassification will only affect the quality of the estimates if there are large inconsistencies between post-strata with highly differential coverage rates.

One must note that inconsistent misclassification is not possible for all A.C.E. postratification variables. Region, metropolitan area size, type of enumeration area, and census mail return rate are all measured at the block level, and are inevitably assigned the same value in both the P and E samples. Inconsistent classification is only possible for the race/ethnicity, owner/renter, and age/sex domains.

We may observe any differences in post-stratification for those people matched between the A.C.E. and the initial census. By assuming that these patterns apply equally to missed people and by working with the observed (estimated) coverage rates, it should be possible to assess the impact of these inconsistencies on the coverage estimates. Of course, this analysis must take into account both the directly reported characteristics and the imputed characteristics, in both the initital census and the A.C.E.

## 4.2 Synthesizing A.C.E. Quality.

### 4.2.1 *Combining the Components of A.C.E. Quality to Assess Accuracy.*

*Analysis Reports Important to this Section:*

- Document 13: "Comparing Accuracy"

One of the larger issues facing the ESCAP Committee will be the problem of summarizing the individual components of error in the A.C.E. into an overall measure of quality. Some errors can be expected to compound each other, for example, false non-matches and failure to identify census fictitious cases, while other errors may offset each other. For example, the error of failure to consistently report Census Day address may offset error introduced by correlation bias.

One method to synthesize error has been the construction of an overall, or total, error model (See Hogan and Wolter, 1988, Mulry and Spencer 1991, 1993, Spencer 2000.) We are working to develop suitable models to synthesize the errors in the 2000 A.C.E. This is a challenging task because many of the errors will not be measured directly. Often we will only have indications of the relative worsening or improvement in the source since 1990, or the assumption that the level remains similar. We will have even less information on any changes in the pattern of the sources of errors, including whether these changes are more defuse or more concentrated. Finally, we will have only a short time between when the information becomes available and when the Director must make the decision. However, given that a total error model could help inform the ESCAP Committee and the Director, we are devoting Bureau resources to this problem and are in the process of obtaining additional resources from a contractor with expertise in this area.

#### 4.2.2 *Comparing the Accuracy of the A.C.E. to the Accuracy of the Uncorrected Census.*

*Analysis Reports Important to this Section:*

- Document 13: "Comparing Accuracy"

Knowing the level of error in the A.C.E. is not enough because the A.C.E. decision will not be made in a vacuum; rather the A.C.E. will be compared to the unadjusted census to determine which is more accurate for redistricting purposes. Both the adjusted and the unadjusted data sets will have their own patterns of error.

As discussed at length in the June, 2000 "Accuracy and Coverage Evaluation: Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000," there are several important criteria in assessing accuracy. For purposes of the ESCAP decision, the Census Bureau will evaluate both numeric and distributive accuracy. Both types of accuracy are important criteria for numbers that will be used in the redistricting process, and both types of accuracy have independent importance as tools in assessing A.C.E. and census quality. Additionally, as discussed in the above document, accuracy can be measured at different geographic levels. The attached analysis reports vary in the geographic level of analysis and the ESCAP Committee will consider these different levels in making its recommendation to the Director.

One simple analysis is to compare estimated undercount rates from the A.C.E. with estimated confidence levels. If, for example, only 50 or so of the 448 post-strata in the A.C.E. measured undercounts at a reasonable level of confidence, one might question whether the A.C.E. had effectively measured any real undercounts. Another perhaps easier to interpret result would be to compare the undercounts among the 64 post-strata groups (collapsed over age and sex) with their confidence intervals. Of course, care must be taken in this analysis, with proper correction for multiple comparisons.

However, if significant undercounts were confirmed with outside information, we would have more faith in those results. For example, if the pattern of "significant" undercounts was confirmed by the results of Demographic Analysis, the demographic estimates program or the full count review program, we would feel more confident in the accuracy of the results. Similarly, we can also look at the reasonableness and consistency of the implied sex ratios of the A.C.E., for example by examining whether the A.C.E. sex rations for ages 18-49 (the prime ages for correlation bias individuals) improve upon the results of the initial census across tenure groups, type of enumeration area, and mail response categories.

One comparison that informed the 1990 decision was a comparison of the measured undercounts for each of the states with the its estimated confidence interval. This evaluation was recorded graphically and was instructive. We intend to repeat this analysis.

Another way to measure overall accuracy is to prepare Loss Functions. Mean squared error is a form of loss function. The Census Bureau prepared Loss Function Analyses in connection with the 1990 adjustment decision, and also in connection with the 1993 decision regarding use of adjusted data as a base for the intercensal estimates. These Loss Functions were able to account for estimated bias in the PES estimates. To apply a loss function, one must decide on one or more levels of analysis. The accuracy criteria discussed above will guide these decisions. For example, one might run Loss Functions to determine the comparative accuracy of two data sets at the state or Congressional district levels. Loss Functions can measure accuracy for both the estimated number of people in each area (numeric accuracy) and the estimated proportion of the population in each area (distributive accuracy).

The 1990 studies and subsequent analyses addressed this issue though complex simulation procedures (See, P-16, as well as Mulry and Spencer [1993]). The Census Bureau concluded that adjustment of the 1990 census would have improved distributive accuracy for states and for areas with populations of more than 100,000. Later Census Bureau work revealed that in general one could not distinguish an improvement in distributive sub-state accuracy for areas with populations of less than 100,000 (Obenski and Fay, 2000).

The problem in constructing loss functions is that the Census Bureau has no external measure of truth at the post-strata and geographic levels (Demographic Analysis provides a benchmark at the national level). Our best measure of the accuracy of the census is the A.C.E. results, corrected by our best measures of A.C.E. error. Because of variance, however, the Census Bureau's best estimate of truth would tend to be more correlated with the A.C.E. than with the census. That is, this estimate would overstate A.C.E. relative accuracy.

The Loss Function Analyses that we hope will be available to inform the ESCAP decision should not be considered determinative for several reasons. First, the Census Bureau will most certainly not complete most of A.C.E. Loss Functions until after the April 1, 2001 deadline. Loss Function Analysis cannot be run until the very end of the A.C.E. process because it requires data that are not available until very late in the process. Second, although A.C.E.

variances will be available, complete information on A.C.E. biases will not be. Accurate bias data are a vital component of any Loss Function. For the purpose of ascertaining preliminary Loss Function information to guide the ESCAP decision, therefore, the Census Bureau will assume that the bias in the A.C.E. is similar to biases in the 1990 PES. If possible, the PES biases will be modified based on an analysis of differences in the PES and the A.C.E., but one should not assume that time will be available to perform this extended research. Finally, one should keep in mind that more complete Loss Functions will be prepared as part of the final evaluation process, many months after the ESCAP decision. These more complete Loss Functions, performed after all the data are available, may well reach results different from those of the preliminary Loss Functions.

The preliminary Loss Functions that will inform the ESCAP decision are still very much in the design phase and cannot be fully described here. In basic terms, however, the Loss Functions will assess numeric and distributive accuracy at the Congressional district and state levels, and probably also for selected sub-state levels.

In view of the limitations on the use of Loss Function Analysis, we wish to stress that Loss Function Analysis will not play a determinative role in the ESCAP Committee's recommendation. The loss functions that will be available to inform the ESCAP Committee will be preliminary, brief, and subject to change with additional data. Loss Functions are important and we will not ignore them; however, the results of the Loss Function Analysis will not be determinative.

## 5. Assessing the Service Based Enumeration Estimation Results

*Analysis Reports Important to this Section:*

- Document 15: "Contributions of Service Based Enumeration Multiplicity Estimation to Corrected Census Results"

The A.C.E. is not the only statistical process that the Census Bureau is preparing to incorporate into the non-apportionment census files. The Census Bureau is preparing estimates of the Service Based Enumeration ("SBE") population based on a multiplicity estimator. The Service Based Enumeration population are those individuals who use homeless shelters, soup-kitchens, and other service providers, individuals who are unlikely to have been included in our usual household or group quarters population.

Because of the transient nature of the SBE population, it is difficult to secure a full count after only a single visit during the initial census. To achieve a better measure of the SBE population, the Census Bureau has asked these individuals which kinds of facilities they use and how often they use the facilities. Census Bureau personnel will then prepare a multiplicity estimator,

using, essentially, a weight equal to the inverse of the frequency of use. For example, a record for a person who used homeless shelters only once during the past week would be given a weight of seven to account for the other six people in a similar situation.

The Census Bureau will, of course, review this estimation program before incorporating its results into the census files.

## 6. Forming an Overall Assessment.

The ESCAP Committee should not anticipate an easy decision. It may be that all the information available to the Committee at the time that it must make its recommendation to the Director will lead clearly in one direction. It is more likely, of course, that the data in the analysis reports will be nuanced. Some of the analysis reports may indicate a high quality A.C.E., while others may raise questions about the quality of the survey results. Some A.C.E. results may be consistent with the preliminary results of Demographic Analysis, while others may raise troubling questions. Similarly, the evidence may clearly infer the conclusion that one data set is superior for certain aggregations, while another data set might be preferable for other aggregations. The immense task of the ESCAP will be to consider the entirety of the evidence before it and carefully assess the relative merits of each data set.

The ESCAP will not make a mechanical decision. It will not depend upon a single hypothesis test or a single measure. Rather, as professionals entrusted with the decision, the ESCAP will assess the totality of the evidence. Taken as a whole, does the evidence point to the conclusion that the A.C.E., while a noble effort, failed to achieve its goals of improving census accuracy and reducing the historic differential undercount. Or, taken as a whole, does the evidence point to the conclusion that using the A.C.E. results would put improved data at our nation's disposal. This synthesis, along with the supporting analysis reports, will attempt to provide the ESCAP with the information it will need to make this immense decision.

## Bibliography


The 1990 "D"Studies, generally.

The 1990 "P" Studies, generally, and specifically the following:

- P-1, "Analysis of Reasonable Imputation Alternatives"
- P-3, "Evaluation of Imputation Methodology for Unresolved Match Status Cases"
- P-4, "Address Misreporting"
- P-5, "Analysis of P-Sample Fabrications from PES Quality Control Data"
- P-5a, "Analysis of Fabrications from Evaluation Follow-up Data"
- P-6, "Fabrication in the P-sample - Interviewer Effect"
- P-7, "Estimates of P-Sample Clerical Matching Error from a Rematching Evaluation"
- P-10, "Measurement of the Census Erroneous Enumeration Clerical Error made in the Assignment of Enumeration Status."
- P-11, "Balancing Error Evaluation"
- P-13, "Use of Alternative Dual System Estimators to Measure Correlation Bias"
- P-16, "Total Error in PES Estimates for Evaluation Post Strata"

Bell, "Using Information from Demographic Analysis in Post-Enumeration Survey Estimation," (1993).

Census Bureau, "Accuracy and Coverage Evaluation: Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000 (2000).

Census Bureau, "Report to Congress – The Plan for Census 2000" (1997).

Cochran, "Sampling Techniques," (1963).

Darga, "Sampling and the Census" (1999).

Fay, R.E., Passel, J.S. Robinson, J.G., and Cowan, D.D. "The Coverage of Population in the 1980 Census," (1988).

Griffiths, R. "Results from the 1995 Census Test: The Contamination Study" (1996).

Himes, C.D., and Clogg, Clifford, "An Overview of Demographic Analysis as a Method for Evaluating Census Coverage in the United States," (1992).

Hogan, H. and Wolter, K.M., "Measuring Accuracy in a Post-Enumeration Survey," (1988).

Mulry, M.H., and Spencer, B.D., "Total Error in PES Estimates of Population," (1991).

Mulry, M.H., and Spencer, B.D., "Accuracy of the 1990 Census and Undercount Adjustment," (1993).

National Research Council, "Modernizing the U.S. Census" (1995)

Obenski, S., and Fay, R.E., "Analysis of CAPE Findings on PES Accuracy at Various Geographic Levels," (2000).

Robinson, J.G., Ahmed, B., Gupta, P.D., and Woodrow, K.A., "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," (1993).

Spencer, B.D., "Adaption of CAPE Loss Function Analysis for Census 2000," (2000) (Draft)